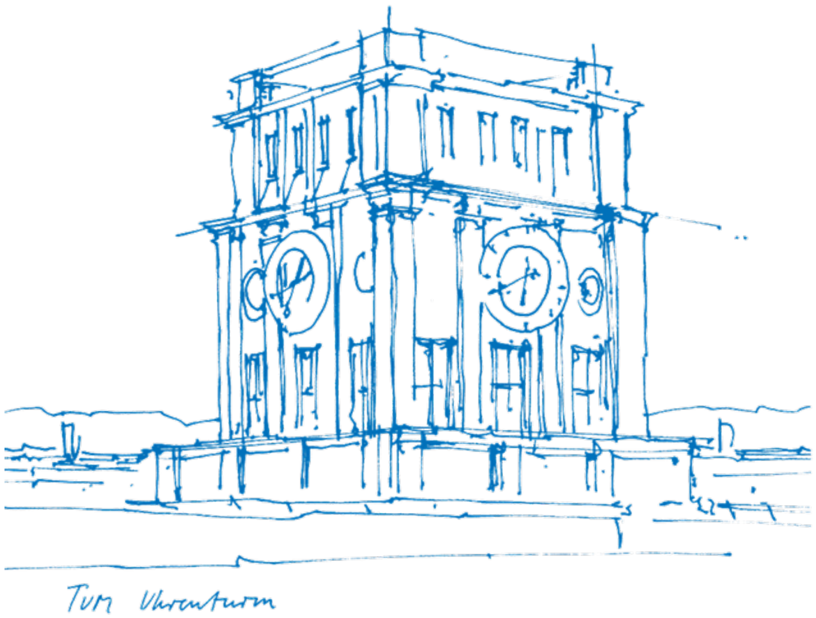
Computer Vision & Artificial Intelligence
Department of Informatics
Technical University of Munich

I : Recap on (Discrete) Probability Theory

Tao Wu, Yuesong Shen, Zhenzhang Ye

Computer Vision & Artificial Intelligence
Technical University of Munich



Probability Space

A **probability space** is a triplet $(\Omega, \mathcal{F}, P)$ consisting of:

- a **sample/state space** Ω — an arbitrary non-empty set.
- a σ**-algebra** — a set of subsets (called events) of Ω s.t.
  - $\Omega \in \mathcal{F}$.
  - $\mathcal{F}$ is closed under complements:
$$A \in \mathcal{F} \;\Rightarrow\; \Omega \backslash A \in \mathcal{F}.$$
  - $\mathcal{F}$ is closed under countable unions:
$$(A_i)_{i\in\mathbb{N}} \in \mathcal{F} \;\Rightarrow\; \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}.$$
- a **probability measure** $P : \mathcal{F} \to [0, 1]$ s.t.
  - P is σ-additive: if $(A_i)_{i\in\mathbb{N}} \in \Omega$ is a countable collection of *pairwise disjoint* sets, then
$$P(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} P(A_i).$$
  - $P(\Omega) = 1$.

Immediate Properties (as exercises)

- $\emptyset \in \mathcal{F}$ and $P(\emptyset) = 0$.
- $A \subset B \ \Rightarrow \ P(A) \leq P(B)$.
- $P(A \cap B) \leq \min(P(A), P(B))$.
- $P(A \cup B) \leq P(A) + P(B)$.
- $P(\Omega \backslash A) = 1 - P(A)$.
- If $(A_i)_{i \in \mathbb{N}} \in \mathcal{F}$ is set of pairwise disjoint events s.t. $\bigcup_{i \in \mathbb{N}} A_i = \Omega$, then $\sum_{i \in \mathbb{N}} P(A_i) = 1$.

Conditional Probability

Assume $P(B) \neq 0$. The **conditional probability** of event A given B is defined as

$$P(A|B) := \frac{P(A \cap B)}{P(B)}. \qquad (\dagger)$$

- Chain rule:

  $$P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1) \ldots P(A_k|A_{k-1} \cap \ldots \cap A_1).$$

  Proof: Recursively apply ($\dagger$).

- Independence: Two events A and B are **independent** ($A \perp B$) iff

  $$P(A \cap B) = P(A)P(B) \;\Leftrightarrow\; P(A|B) = P(A).$$

Discrete vs. Continuous Probability Space

- Discrete case (sufficient for this course):
  - Ω contains at most countably many elements.
  - If Ω is finite, we typically set $\mathcal{F} = 2^\Omega$ (2^Ω stands for the power set of Ω).
  - P is characterized by **probability mass function** $p : \Omega \to [0, 1]$ s.t. $\sum_{\omega \in \Omega} p(\omega) = 1$. This means:

$$\forall A \in \mathcal{F} : \; P(A) = \sum_{\omega \in A} p(\omega).$$

- Continuous case:
  - Ω is a (possibly uncountable) *measurable space*.
  - P is a *probability measure*.
  - P often admits a **probability density function** $p : \Omega \to [0, 1]$ ($\Omega \subset \mathbb{R}^n$), i.e.,

$$\forall A \in \mathcal{F} : \; P(A) = \int_A p(\omega) d\omega.$$

  Formally, $p = \frac{dP}{d\omega}$ is the *Radon-Nikodym derivative* of P w.r.t. the Lebesgue measure.

Random Variable

A **random variable** X on $(\Omega, \mathcal{F}, P)$ is a *measurable function* $X : \Omega \to \mathcal{X}$ s.t.

$$\forall A \subset \mathcal{X} \text{ measurable}: \; P(X \in A) := P(X^{-1}(A)) = P(\{\omega \in \Omega : X(\omega) \in A\}).$$

For discrete RV, it's convenient to directly work on the (discrete) output space $\mathcal{X}$:

$$p(x) := P(X = x) = \sum_{\omega : X(\omega) = x} P(\omega).$$

- **Expectation**:

$$\begin{aligned} \mathbb{E}[X] &:= \int_\Omega X(\omega) dP(\omega) && \text{(continuous)} \\ &= \sum_{\omega \in \Omega} X(\omega) p(\omega) = \sum_x x p(x). && \text{(discrete)} \end{aligned}$$

- **Covariance** (X, Y are both real-valued RV):

$$\begin{aligned} \text{Cov}[X, Y] &:= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \end{aligned}$$

- **Variance**: $\text{Var}[X] := \text{Cov}[X, X]$.

Simple Properties (as exercises)

Expectation:

- $\mathbb{E}[a] = a$ for any constant a.
- Linearity:

$$\begin{aligned} \mathbb{E}[X+Y] &= \mathbb{E}[X] + \mathbb{E}[Y], \\ \mathbb{E}[\alpha X] &= \alpha \mathbb{E}[X]. \quad (\alpha \text{ is a constant scalar}) \end{aligned}$$

Covariance:

- $\mathrm{Cov}[X, a] = \mathrm{Cov}[a, X] = 0$ for any constant a.
- Homogeneity:

$$\mathrm{Cov}[\alpha X, \beta Y] = \alpha\beta\, \mathrm{Cov}[X, Y]. \quad (\alpha, \beta \text{ are constant scalars})$$

- Covariance under independence $X \perp Y$:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \quad \Leftrightarrow \quad \mathrm{Cov}[X, Y] = 0.$$

Joint, Marginal, and Conditional Probability

Let two discrete RVs X, Y be given.

Joint probability:

$$p(x, y) = P(X = x, Y = y).$$

Marginal probability:

$$p(x) = \sum_y p(x, y).$$

Conditional probability:

$$p(x|y) = P(X = x|Y = y) = \frac{p(x, y)}{p(y)} = \frac{p(x, y)}{\sum_x p(x, y)}.$$

Conditional independence between X, Y given Z (i.e. $X \perp Y \mid Z$) iff

$$p(x, y|z) = p(x|z)p(y|z) \;\Leftrightarrow\; p(x|y, z) = p(x|z).$$

Bayes' rule:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}.$$

Computer Vision & Artificial Intelligence
Department of Informatics
Technical University of Munich

Example on Binary Segmentation

Consider a (simplified) binary segmentation problem:
$X \in \{\text{Dark}, \text{Bright}\}, \; Y \in \{\text{Foreground}, \text{Background}\}$.

$p(x, y)$	Dark	Bright	$p(y)$
Foreground	0.163	0.006	0.169
Background	0.116	0.715	0.831
$p(x)$	0.279	0.721	

Figure: Probability Table for Binary Segmentation.

- Conditional probability via joint probability:

$$P(Y = \text{Foregr.} | X = \text{Dark}) = \frac{P(X = \text{Dark}, Y = \text{Foregr.})}{P(X = \text{Dark})} = \frac{0.163}{0.279} = 0.584.$$

- Conditional probability via Bayes' rule:

$$P(X = \text{Dark} | Y = \text{Foregr.}) = \frac{P(Y = \text{Foregr.} | X = \text{Dark}) P(X = \text{Dark})}{P(Y = \text{Foregr.})} = \frac{0.584 \cdot 0.279}{0.169} = 0.964.$$